# Note

Feng-Yang Hsieh

## 1 Sample

The vector boson samples are generated from MadGraph5 aMC@NLO 3.3.1. Then pass to Pythia for showering and hadronization. Then pass to Delphes for detector simulation.

The processes are the VBF production for doubly charged Higgses $H_5^{\pm\pm}$ and heavy neutral Higgses $H_5$ with decays $H_5^{\pm\pm} \to W^{\pm}W^{\pm} \to jjjj$ and $H_5 \to ZZ \to jjjj$.

Following are the MadGraph scripts for generating $W^+$ sample:

```
import model GM_NLO
define v = z w+ w-
generate p p > H5pp j j $$v, (H5pp > w+ w+, (w+ > j j), (w+ > j j))
output VBF_H5pp_ww_jjjj
launch VBF_H5pp_ww_jjjj

shower=Pythia8
detector=Delphes
analysis=OFF
done

set param_card tanth 0.226480
set param_card lam2 0.070070
set param_card lam3 -1.331328
set param_card lam4 1.364671
set param_card lam5 -1.963271
set param_card M1coeff 1046.827111
set param_card M2coeff 135.30791
```

```
set run_card nevents 100000
set run_card ebeam1 6500.0
set run_card ebeam2 6500.0
set run_card cut_decays True
set run_card ptj 50
set run_card etaj 3

/home/r10222035/boosted_V_ML_test/Cards/delphes_card.dat

done
```

In Delphes, jets are constructed by the anti-$k_t$ algorithm with $R = 0.7$.

## 1.1 Sample selection

Only the events that satisfy the following requirement will use for training.

1. The transverse momentum of jets are required $p_{\mathrm{T}} \in (350, 450)$ GeV and in range $|\eta| < 1$.
2. Merging: The angular distance between the two quarks decayed from the vector boson is required $\Delta R(q_1, q_2) < 0.6$.
3. Matching: The vector boson and jet are matched if $\Delta R(V, j) < 0.1$. The events with less than two matching jets will be discarded.

## 1.2 Sample pre-processing

The jet charge of a jet is defined as

$$\mathcal{Q}_\kappa \equiv \frac{1}{(p_{\mathrm{T},J})^\kappa} \sum_{i \in J} q_i \times \left(p_{\mathrm{T}}^i\right)^\kappa \tag{1}$$

where $i$ represents the constituent in the jet $J$.

### 1.2.1 Jet image

After the sample selection, we can construct the jet image of the matching vector boson jet. The jet image is constructed by the following steps:

1. Centralization: Calculating the $p_{\mathrm{T}}$ weighted center in $\eta, \phi$ plane, then shift this point to origin.

2. Rotation
3. Flipping
4. Pixelating in $\Delta\eta = \Delta\phi = 1.6$ box, with $75 \times 75$ pixels.

## 1.3 Plots

Figure 1 is the jet mass distribution.

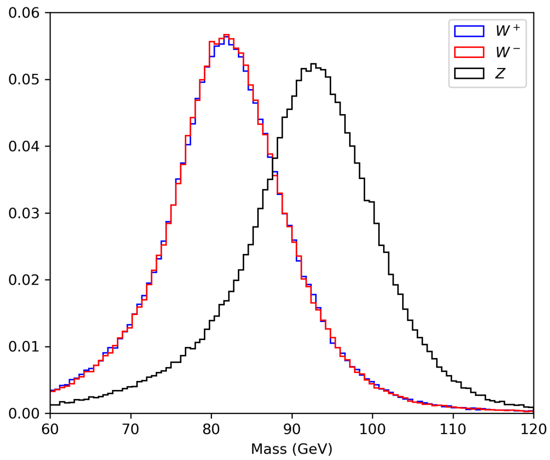



Figure 1: Jet mass distributions of their vector boson samples.

Figure 2 is the jet charge distributions with different $\kappa$.

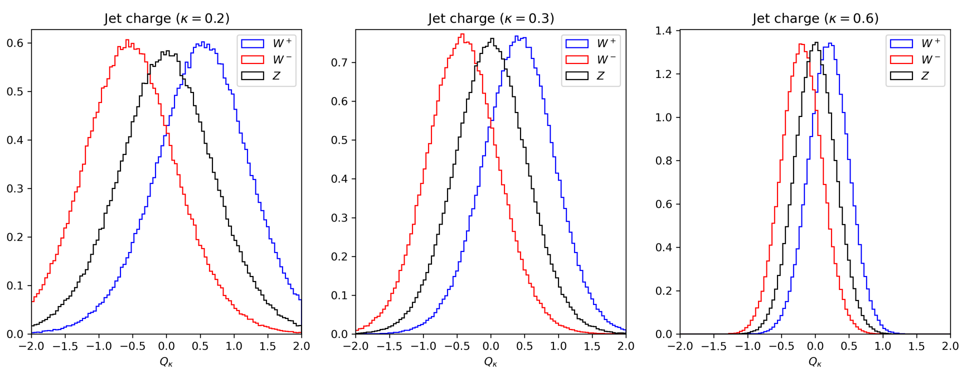



Figure 2: $\mathcal{Q}_\kappa$ distributions of vector boson samples.

Figure 3, 4 are the average jet images of $p_{\mathrm{T}}$ and $\mathcal{Q}_\kappa$. Figure 5 is the $Z$ jet image minus $W^+$ jet image in $p_{\mathrm{T}}$ channel.

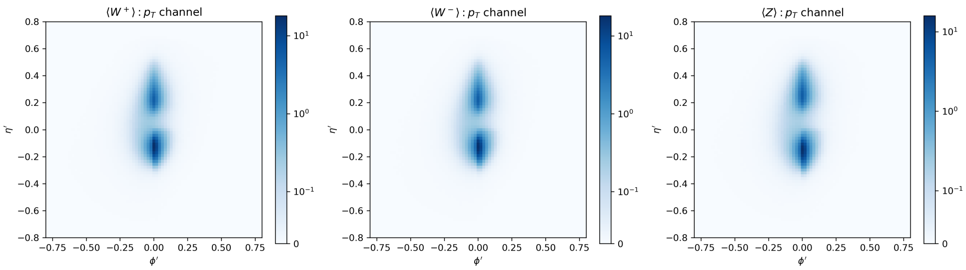



Figure 3: Average of jet images in the $p_{\mathrm{T}}$ channel.

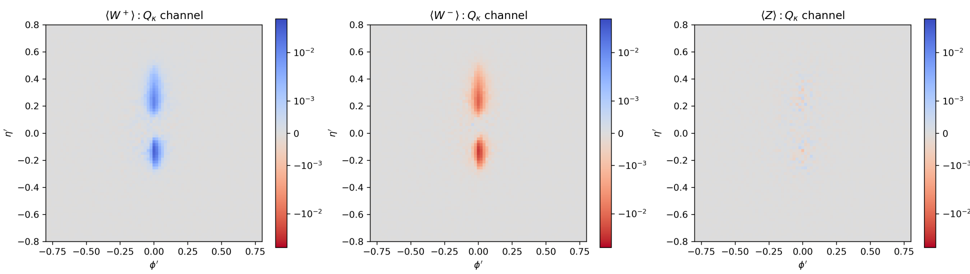



Figure 4: Average of jet images in the $\mathcal{Q}_\kappa$ channel, with $\kappa = 0.15$.

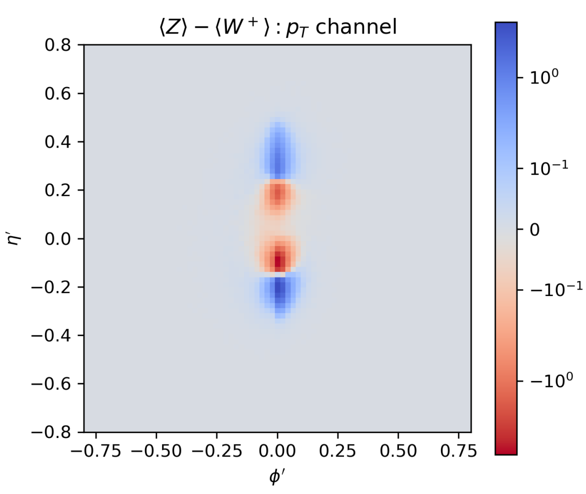



Figure 5: The difference between the $Z$ and $W^+$ average jet images in $p_{\mathrm{T}}$ channel.

# 2 BDT

The input features are jet mass $\mathcal{M}$ and jet charge $\mathcal{Q}_\kappa$.

## 2.1 BDT training

Use the GBDT implemented by scikit-learn. The parameters are set as the default setting in sklearn.

The training and testing sample size are in Table 1.

Table 1: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Testing set |
|---|---|---|
| 1 | $14k + 15k + 13k$ | $3k + 3k + 3k$ |
| 2 | $100k + 105k + 90k$ | $25k + 26k + 22k$ |
| 3 | $242k + 257k + 221k$ | $60k + 64k + 54k$ |

## 2.2 BDT results

The training results are summarized in Table 2. The ROC of BDT for $\kappa = 0.3$ is presented in Figure 6.

Table 2: The training results of BDT for different $\kappa$ samples.

| | | | $W^+$ | | $W^-$ | | $Z$ | |
|---|---|---|---|---|---|---|---|---|
| | Sample | Overall ACC | AUC | ACC | AUC | ACC | AUC | ACC |
| BDT $\kappa = 0.15$ | 1 | 0.644 | 0.808 | 0.773 | 0.812 | 0.766 | 0.812 | 0.770 |
| BDT $\kappa = 0.30$ | | 0.649 | 0.822 | 0.778 | 0.822 | 0.769 | 0.814 | 0.773 |
| BDT $\kappa = 0.15$ | 2 | 0.646 | 0.808 | 0.769 | 0.815 | 0.765 | 0.809 | 0.774 |
| BDT $\kappa = 0.30$ | | 0.652 | 0.820 | 0.774 | 0.826 | 0.772 | 0.810 | 0.775 |
| BDT $\kappa = 0.15$ | 3 | 0.648 | 0.811 | 0.769 | 0.817 | 0.767 | 0.810 | 0.775 |
| BDT $\kappa = 0.30$ | | 0.654 | 0.824 | 0.776 | 0.829 | 0.773 | 0.811 | 0.774 |

# 3 CNN

The inputs are the jet image of $p_T$ and $\mathcal{Q}_\kappa$.

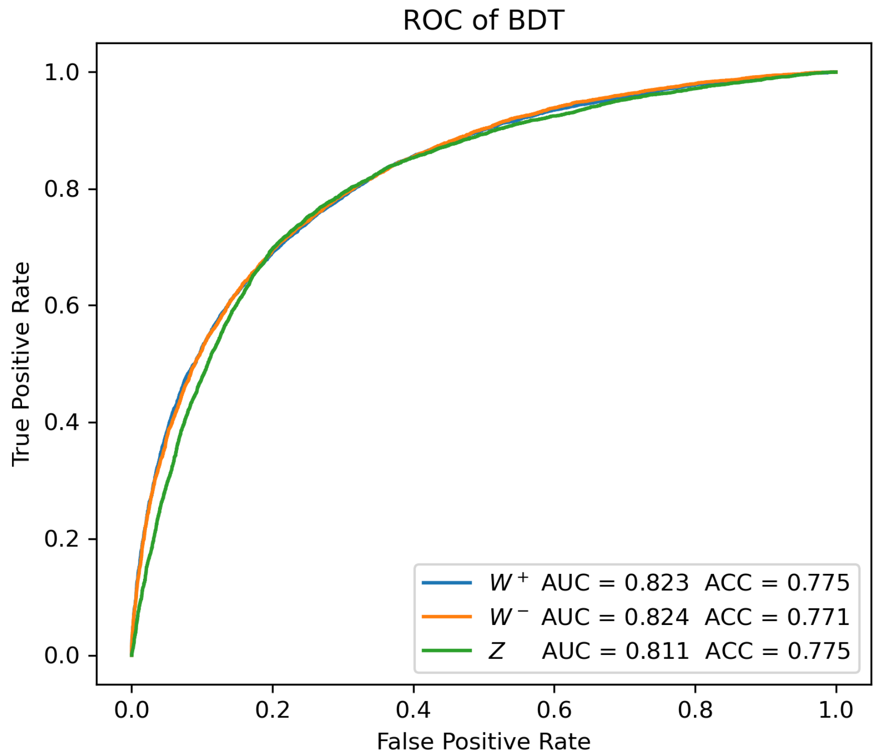



Figure 6: The ROC of BDT. Case $W^+$ means take $W^+$ samples as signal and $W^-, Z$ as background.

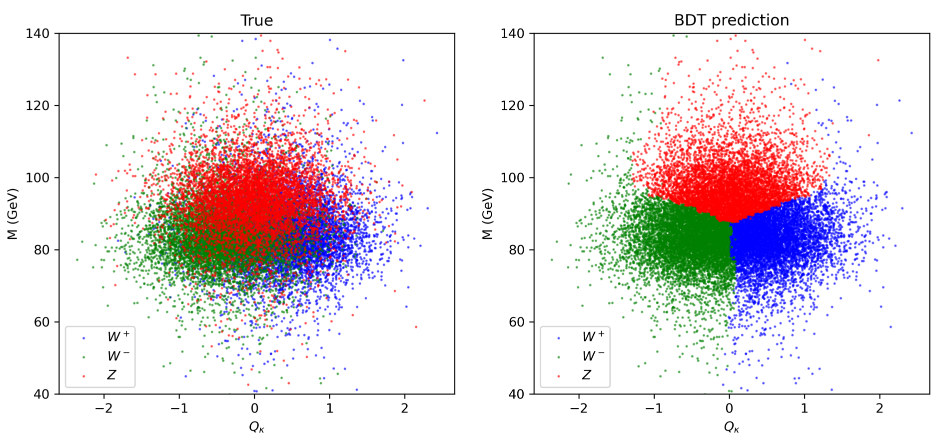



Figure 7: The left plot shows the true distributions for $W^+, W^-, Z$ in the $(\mathcal{Q}_\kappa, \mathcal{M})$ plane (for $\kappa = 0.3$). The right plot is the BDT prediction.

## 3.1 CNN training

The training, validation, and testing sample size are in Table 3.

Table 3: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Validation set | Testing set |
|---|---|---|---|
| 1 | $11k + 12k + 10k$ | $2k + 2k + 2k$ | $3k + 3k + 3k$ |
| 2 | $80k + 84k + 72k$ | $19k + 21k + 18k$ | $25k + 26k + 22k$ |
| 3 | $194k + 205k + 176k$ | $48k + 51k + 44k$ | $60k + 64k + 55k$ |

## 3.2 CNN results

The training results are summarized in Table 4. The ROC of CNN for $\kappa = 0.15$ is presented in Figure 8.

Table 4: The training results of CNN with different $\kappa$ samples.

| | | | $W^+$ | | $W^-$ | | $Z$ | |
|---|---|---|---|---|---|---|---|---|
| | Sample | Overall ACC | AUC | ACC | AUC | ACC | AUC | ACC |
| CNN $\kappa = 0.15$ | 1 | 0.639 | 0.825 | 0.772 | 0.829 | 0.774 | 0.782 | 0.768 |
| CNN $\kappa = 0.30$ | | 0.632 | 0.822 | 0.771 | 0.825 | 0.771 | 0.786 | 0.771 |
| CNN $\kappa = 0.15$ | 2 | 0.667 | 0.847 | 0.790 | 0.847 | 0.784 | 0.828 | 0.795 |
| CNN $\kappa = 0.30$ | | 0.663 | 0.843 | 0.786 | 0.843 | 0.779 | 0.825 | 0.795 |
| CNN $\kappa = 0.15$ | 3 | 0.672 | 0.849 | 0.793 | 0.851 | 0.787 | 0.834 | 0.799 |
| CNN $\kappa = 0.30$ | | 0.666 | 0.845 | 0.788 | 0.847 | 0.783 | 0.832 | 0.800 |

# 4 CNN$^2$

## 4.1 CNN$^2$ training

The training, validation, and testing sample size are in Table 5.

## 4.2 CNN$^2$ results

The training results are summarized in Table 6.

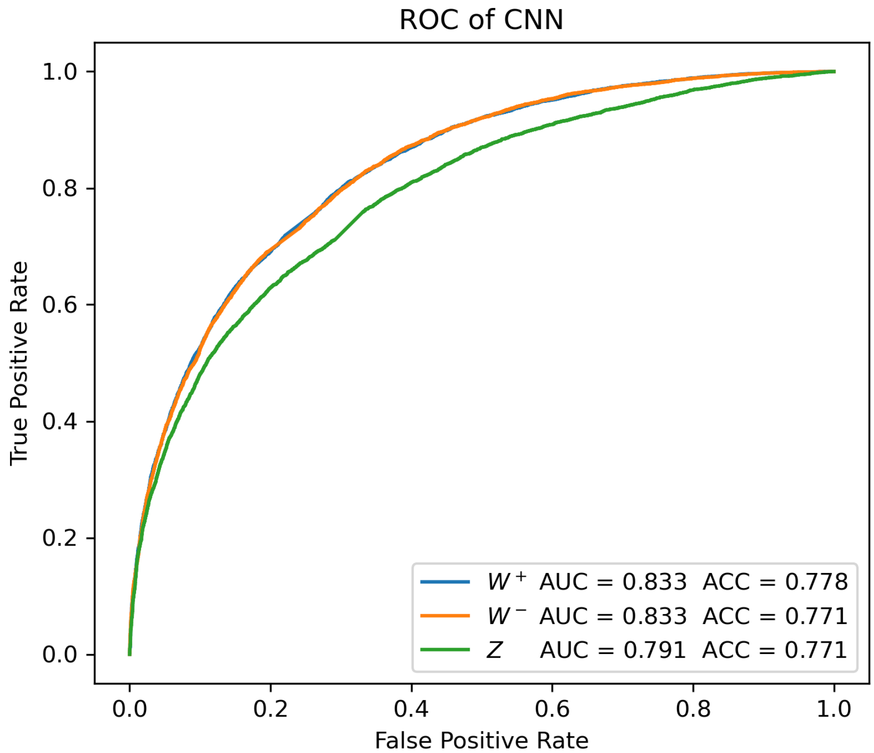



Figure 8: The ROC of CNN for $\kappa = 0.15$. Case $W^+$ means take $W^+$ samples as signal and $W^-, Z$ as background.

Table 5: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Validation set | Testing set |
|---|---|---|---|
| 1 | $11k + 12k + 10k$ | $2k + 2k + 2k$ | $3k + 3k + 3k$ |
| 2 | $80k + 84k + 72k$ | $19k + 21k + 18k$ | $25k + 26k + 22k$ |
| 3 | $194k + 205k + 176k$ | $48k + 51k + 44k$ | $60k + 64k + 55k$ |

Table 6: The training results of $\text{CNN}^2$ for different $\kappa$ samples.

| | Sample | Overall ACC | $W^+$ AUC | $W^+$ ACC | $W^-$ AUC | $W^-$ ACC | $Z$ AUC | $Z$ ACC |
|---|---|---|---|---|---|---|---|---|
| $\text{CNN}^2$ $\kappa = 0.15$ | 1 | 0.642 | 0.828 | 0.773 | 0.831 | 0.773 | 0.801 | 0.780 |
| $\text{CNN}^2$ $\kappa = 0.30$ | | 0.645 | 0.823 | 0.773 | 0.828 | 0.773 | 0.802 | 0.784 |
| $\text{CNN}^2$ $\kappa = 0.15$ | 2 | 0.677 | 0.848 | 0.792 | 0.848 | 0.786 | 0.842 | 0.806 |
| $\text{CNN}^2$ $\kappa = 0.30$ | | 0.669 | 0.842 | 0.788 | 0.843 | 0.780 | 0.839 | 0.804 |
| $\text{CNN}^2$ $\kappa = 0.15$ | 3 | 0.674 | 0.848 | 0.792 | 0.849 | 0.785 | 0.843 | 0.807 |
| $\text{CNN}^2$ $\kappa = 0.30$ | | 0.675 | 0.846 | 0.790 | 0.847 | 0.784 | 0.845 | 0.808 |

# 5 Modify preprocess and preselection

Preprocess: For the vector boson jet which $\phi$ coordinate across the $\phi = \pm\pi$ boundary, plus its $\phi$ coordinate by $\pi$ such that its center is located around $\phi = 0$.

Preselection: Include the events with only one jet that can pass the cuts.

## 5.1 CNN results for modified preprocess and preselection

The training, validation, and testing sample size are in Table 7.

Table 7: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Validation set | Testing set |
|---|---|---|---|
| 1 | $313k + 323k + 307k$ | $78k + 80k + 76k$ | $97k + 100k + 96k$ |

The training results are summarized in Table 8.

Table 8: The training results of CNN with modified preprocessing and preselection.

| | Sample | Overall ACC | $W^+$ AUC | $W^+$ ACC | $W^-$ AUC | $W^-$ ACC | $Z$ AUC | $Z$ ACC |
|---|---|---|---|---|---|---|---|---|
| CNN $\kappa = 0.15$ | 1 | 0.684 | 0.857 | 0.798 | 0.855 | 0.794 | 0.828 | 0.793 |
| $\text{CNN}^2$ $\kappa = 0.15$ | 1 | 0.684 | 0.855 | 0.797 | 0.854 | 0.793 | 0.835 | 0.797 |

## 5.2 CNN results for Fishbone

The training, validation, and testing sample size are in Table 9. For case 1, only consider the events with 2 jets that can pass the cuts. For case 2, the events with only 1 jet can pass the cuts are included. The training results are summarized in Table 10.

Table 9: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Validation set | Testing set |
|---|---|---|---|
| 1 | $61k + 65k + 102k$ | $15k + 16k + 25k$ | $19k + 20k + 32k$ |
| 2 | $313k + 323k + 307k$ | $78k + 81k + 76k$ | $97k + 101k + 95k$ |

Table 10: The training results of CNN with Fishbone's code.

| | | | $W^+$ | | $W^-$ | | $Z$ | |
|---|---|---|---|---|---|---|---|---|
| | Sample | Overall ACC | AUC | ACC | AUC | ACC | AUC | ACC |
| CNN $\kappa = 0.15$ | 1 | 0.694 | 0.857 | 0.820 | 0.856 | 0.814 | 0.842 | 0.776 |
| CNN $\kappa = 0.15$ | 2 | 0.682 | 0.856 | 0.797 | 0.854 | 0.793 | 0.827 | 0.791 |

# 6 Old sample for training

This section uses the model `GM_UFO` and different parameter settings. The model and parameters are the same as the paper. The selection criteria are the same as Sec.1.1. The events with only one vector boson passing the criteria are included.

Following are the MadGraph scripts for generating $W^+$ sample:

```
import model GM_UFO
define v = z w+ w-
generate p p > H5pp j j $$v, (H5pp > w+ w+, (w+ > j j), (w+ > j j))
output VBF_H5pp_ww_jjjj-J

launch VBF_H5pp_ww_jjjj-J

shower=Pythia8
detector=Delphes
analysis=OFF
done

set param_card tanth 2.234400e+01
set param_card lam2 1.040100e+00
```

```
set param_card lam3 8.829540e+00
set param_card lam4 -2.232270e+00
set param_card lam5 7.672600e+00
set param_card M1coeff 1.000000e+02
set param_card M2coeff 1.000000e+02

set param_card wt 1.000000e+00
set param_card wz 1.000000e+00
set param_card ww 1.000000e+00
set param_card wh Auto
set param_card wh__2 Auto
set param_card wh3p Auto
set param_card wh3z Auto
set param_card wh5pp Auto
set param_card wh5p Auto
set param_card wh5z Auto

set run_card nevents 10000
set run_card ebeam1 6500.0
set run_card ebeam2 6500.0

/home/r10222035/boosted_V_ML_test/Cards/delphes_card.dat

done
```

## 6.1 Training results

The training, validation, and testing sample size are in Table 11.

Table 11: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Validation set | Testing set |
|---|---|---|---|
| 1 | $112k + 116k + 101k$ | $27k + 29k + 25k$ | $34k + 36k + 31k$ |
| 2 | $224k + 232k + 201k$ | $56k + 58k + 50k$ | $69k + 72k + 63k$ |

The training results are summarized in Table 12. The results are better than the Sec.5.1.

Table 12: The training results of CNN with modified preprocessing and preselection.

| | Sample | Overall ACC | $W^+$ AUC | $W^+$ ACC | $W^-$ AUC | $W^-$ ACC | $Z$ AUC | $Z$ ACC |
|---|---|---|---|---|---|---|---|---|
| CNN $\kappa = 0.15$ | 1 | 0.690 | 0.861 | 0.798 | 0.858 | 0.793 | 0.830 | 0.808 |
| CNN$^2$ $\kappa = 0.15$ | | 0.693 | 0.860 | 0.797 | 0.859 | 0.794 | 0.838 | 0.816 |
| CNN $\kappa = 0.15$ | 2 | 0.697 | 0.864 | 0.801 | 0.863 | 0.796 | 0.837 | 0.812 |
| CNN$^2$ $\kappa = 0.15$ | | 0.698 | 0.863 | 0.800 | 0.862 | 0.797 | 0.844 | 0.818 |

# 7 Full event sample

For the full event case, there are six processes: $H_5^{\pm\pm} \to W^\pm W^\pm \to jjjj$, $H_5^\pm \to W^\pm Z \to jjjj$, $H_5 \to ZZ \to jjjj$, $H_5 \to W^+W^- \to jjjj$.

## 7.1 Full event sample selection

Only the events that satisfy the following requirements will be used for training.

1. The transverse momentum of vector boson jets are required $p_\mathrm{T} \in (350, 450)$ GeV and in range $|\eta| < 1$.

2. Merging: The angular distance between the two quarks decayed from the vector boson is required $\Delta R(q_1, q_2) < 0.6$.

3. Matching: The vector boson and jet are matched if $\Delta R(V, j) < 0.1$. The events with less than two matching jets will be discarded.

## 7.2 Full event sample pre-processing

### 7.2.1 Event image

After the sample selection, we can construct the event image. The event image contains 2 vector boson jets and 2 forward jets. The vector boson jets are selected by the merging requirement. The forward jets are selected by the highest $p_\mathrm{T}$ jets from the remaining jets. The event image is constructed by the following steps:

1. Centralization: Calculating the $p_\mathrm{T}$ weighted center in $\phi$ coordinate, then shift this point to origin.

2. Flipping: Flip the highest intensity quadrant to the first quadrant.

3. Pixelating in $\Delta\eta = \Delta\phi = 6$ box, with $75 \times 75$ pixels.

### 7.3 Event sample plots

Figure 9, 10 are the single event images of $p_{\mathrm{T}}$ and $\mathcal{Q}_\kappa$. Figure 11, 12 are the average event images of $p_{\mathrm{T}}$ and $\mathcal{Q}_\kappa$.

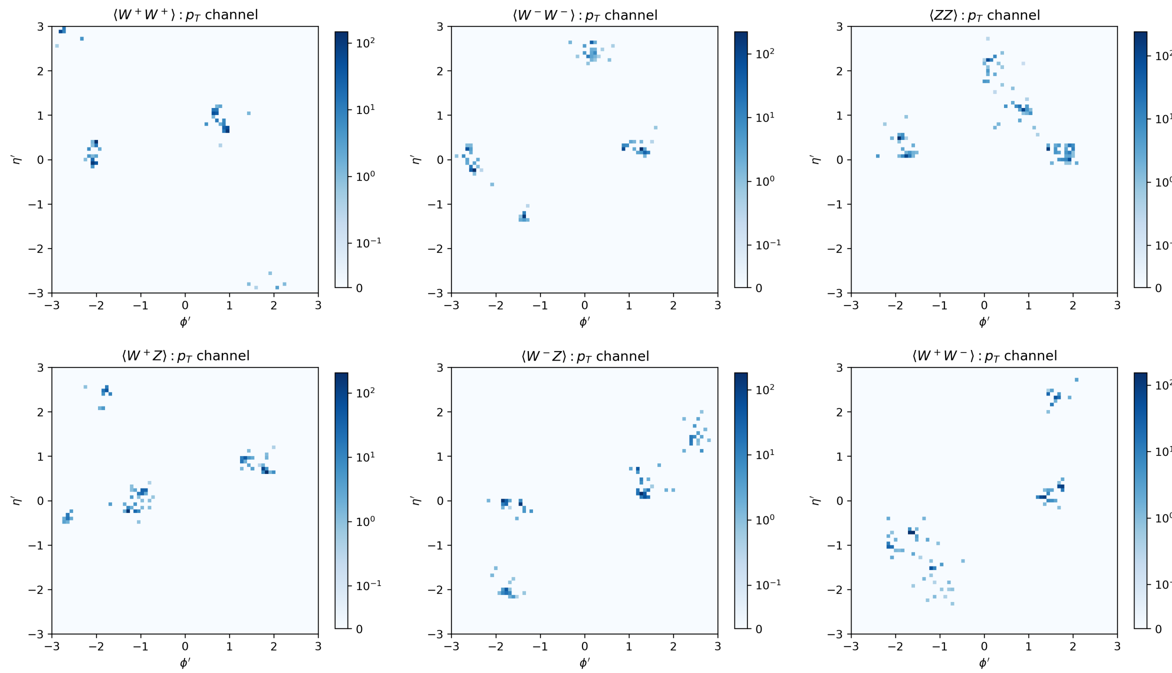



Figure 9: Single event images in the $p_{\mathrm{T}}$ channel.

Figure 13 is the $ZZ$ event image minus $W^+W^+$ event image in $p_{\mathrm{T}}$ channel.

## 8 Correct decay width sample

In this section, the samples are generated with the correct decay widths, i.e., the decay width of $t, W$, and $Z$ do not change and the exotic Higgses in the GM model are set to "Auto" , meaning the decay widths are calculated by MadGraph.

Following are the MadGraph scripts for generating $W^+$ sample:

```
import model GM_UFO
define v = z w+ w-
generate p p > H5pp j j $$v, (H5pp > w+ w+, (w+ > j j), (w+ > j j))
output VBF_H5pp_ww_jjjj

launch VBF_H5pp_ww_jjjj
```

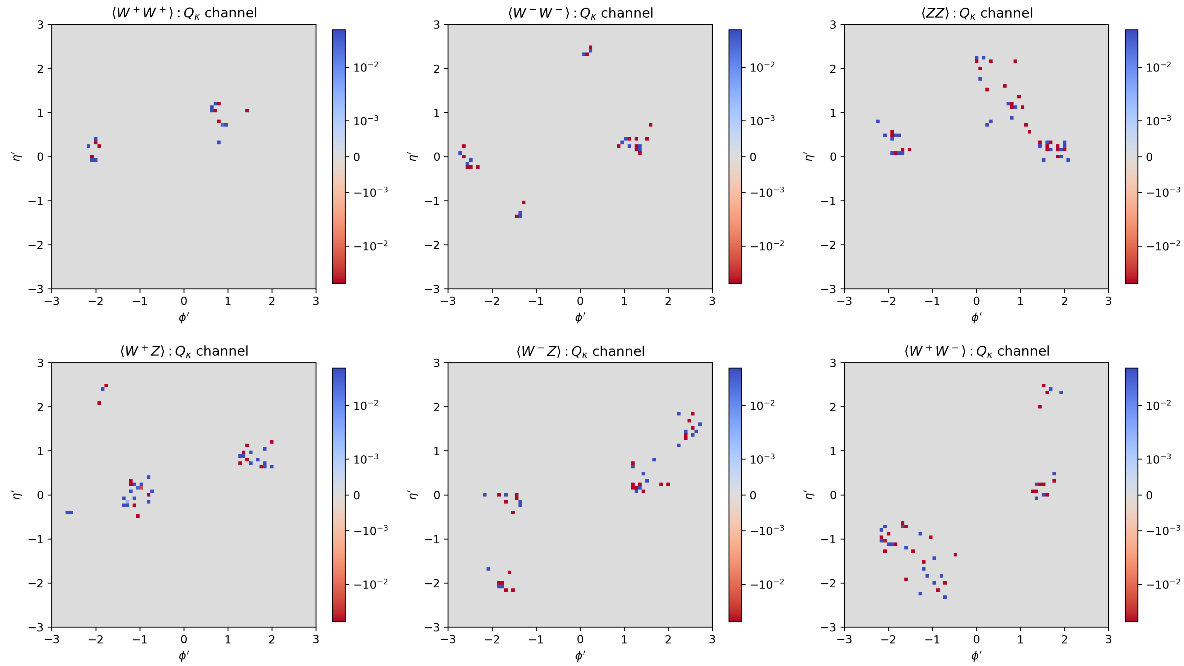



Figure 10: Single event images in the $\mathcal{Q}_\kappa$ channel, with $\kappa = 0.15$.

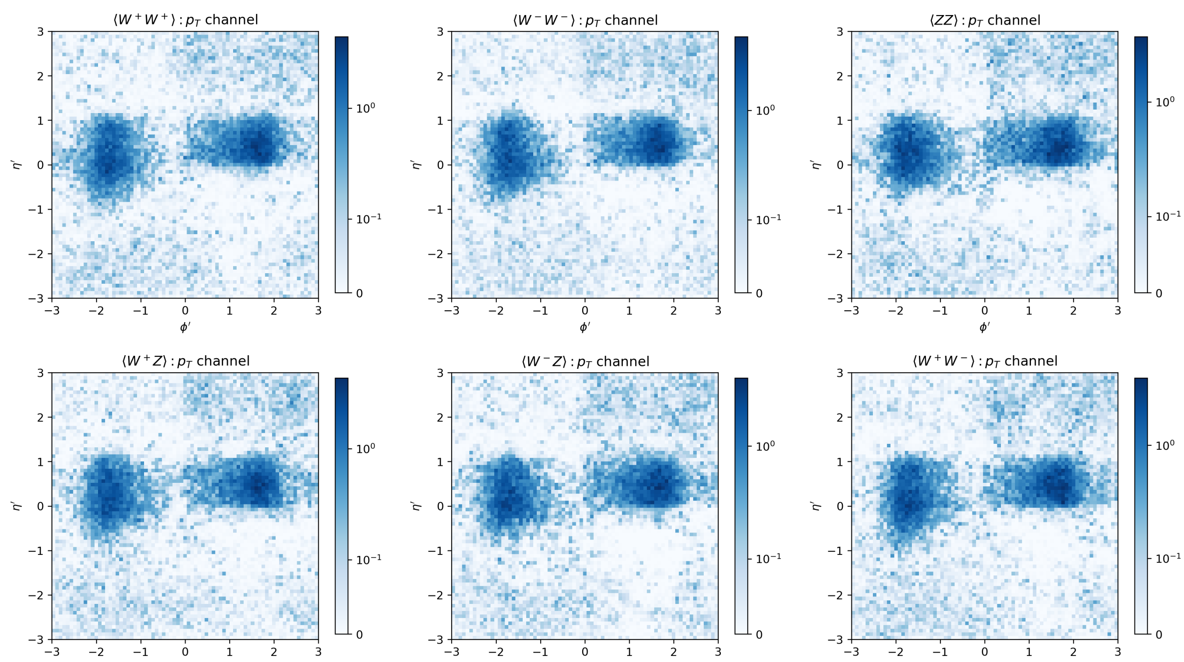



Figure 11: Average of event images in the $p_{\mathrm{T}}$ channel.

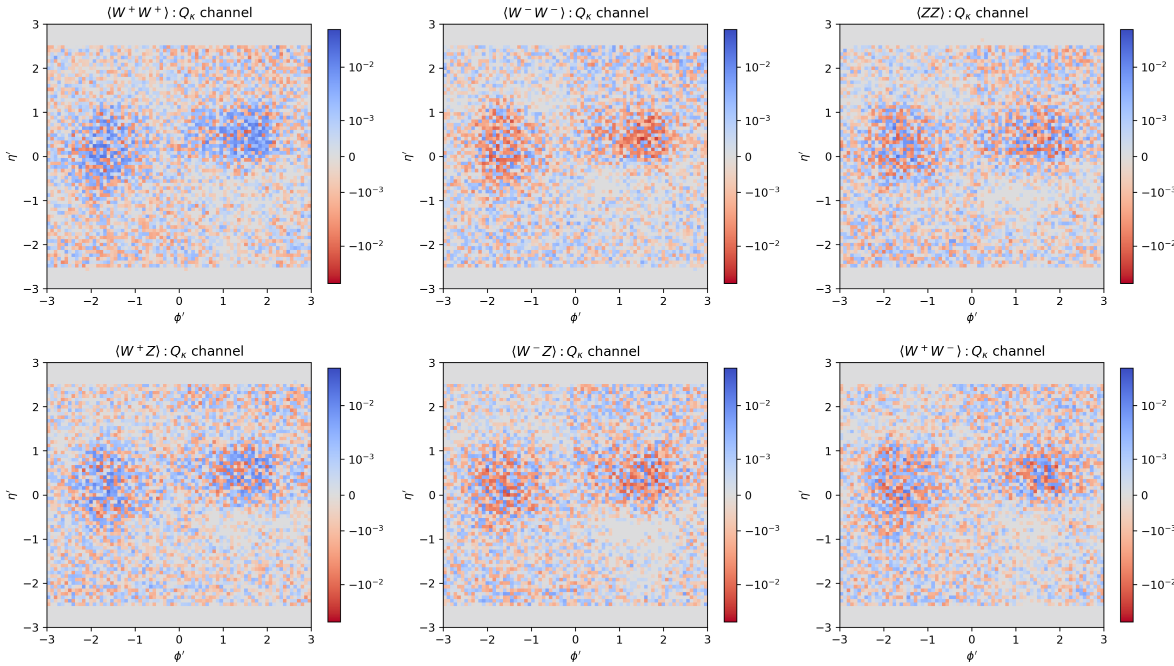



Figure 12: Average of event images in the $\mathcal{Q}_\kappa$ channel, with $\kappa = 0.15$.

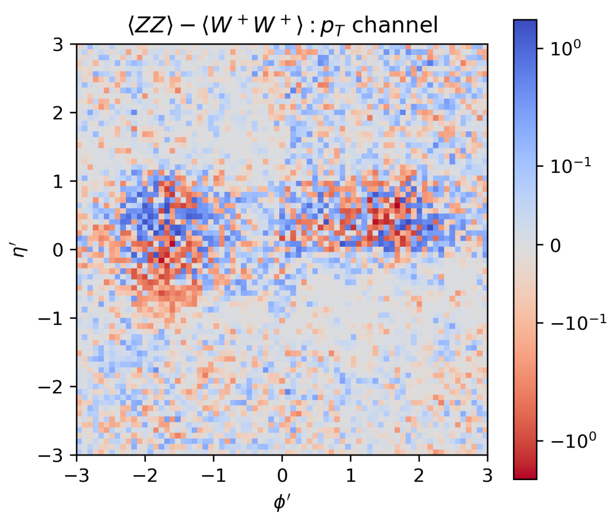



Figure 13: The difference between the $ZZ$ and $W^+W^+$ average event images in $p_{\mathrm{T}}$ channel.

```
shower=Pythia8
detector=Delphes
analysis=OFF
done

set param_card tanth 2.234400e+01
set param_card lam2 1.040100e+00
set param_card lam3 8.829540e+00
set param_card lam4 -2.232270e+00
set param_card lam5 7.672600e+00
set param_card M1coeff 1.000000e+02
set param_card M2coeff 1.000000e+02

set param_card wh Auto
set param_card wh__2 Auto
set param_card wh3p Auto
set param_card wh3z Auto
set param_card wh5pp Auto
set param_card wh5p Auto
set param_card wh5z Auto

set run_card nevents 10000
set run_card ebeam1 6500.0
set run_card ebeam2 6500.0

/home/r10222035/boosted_V_ML_test/Cards/delphes_card.dat

done
```

## 8.1 Training results

The training, validation, and testing sample size are in Table 13.

Table 13: The entries in the sum correspond to the $(W^+, W^-, Z)$ samples.

| Case | Training set | Validation set | Testing set |
|---|---|---|---|
| 1 | $223k + 232k + 202k$ | $56k + 57k + 50k$ | $69k + 72k + 62k$ |

The training results are summarized in Table 14. The results are similar to the Sec.5.1

Table 14: The training results of correct decay width sample. The training results of CNN and CNN$^2$ are presented with an average and a standard deviation. These values are derived from 10 times training with the same dataset. Yet, the ACC value of each boson is only derived from a single result.

| | | | $W^+$ | | $W^-$ | | $Z$ | |
|---|---|---|---|---|---|---|---|---|
| | Sample | Overall ACC | AUC | ACC | AUC | ACC | AUC | ACC |
| BDT $\kappa = 0.30$ | | 65.1 | 82.6 | 77.5 | 82.5 | 77.0 | 79.2 | 77.4 |
| CNN $\kappa = 0.15$ | 1 | 68.24±0.04 | 85.72±0.02 | (79.5) | 85.58±0.02 | (79.1) | 81.64±0.05 | (79.8) |
| CNN$^2$ $\kappa = 0.15$ | | 68.26±0.13 | 85.55±0.04 | (79.5) | 85.41±0.05 | (79.0) | 82.22±0.08 | (80.2) |

but worse than Sec.6.1. It seems that the decay width of $t, W$, and $Z$ will affect the training results.

Figure 14 is CNN's loss and accuracy curve. Figure 15 is CNN's ROC.

Figure 16 is CNN$^2$'s loss and accuracy curve. Figure 17 is CNN$^2$'s ROC.

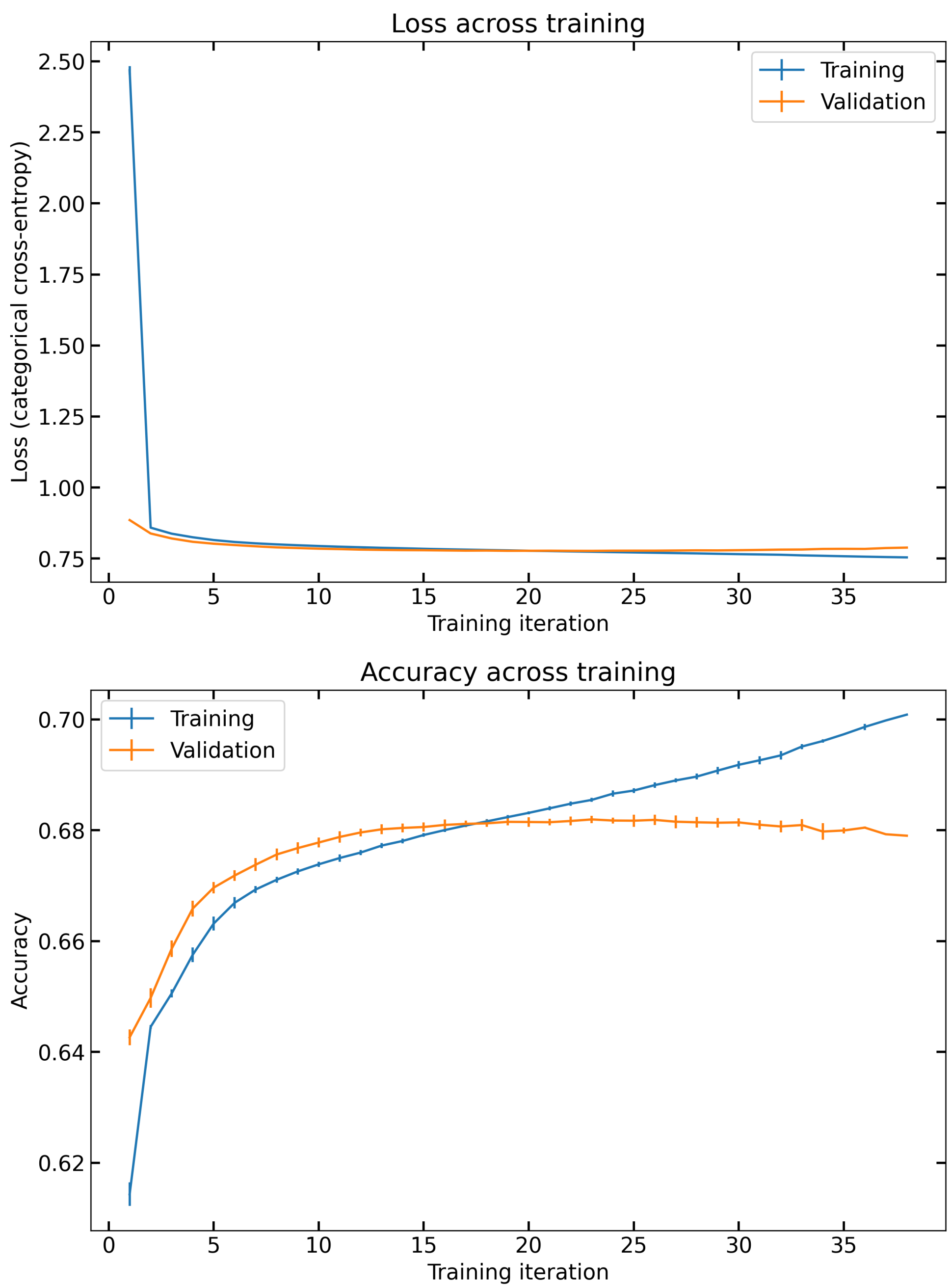



Figure 14: The loss and accuracy curve for the CNN model. Both of them are demonstrated with the average value (solid curve) and the first standard deviation range (error bar).

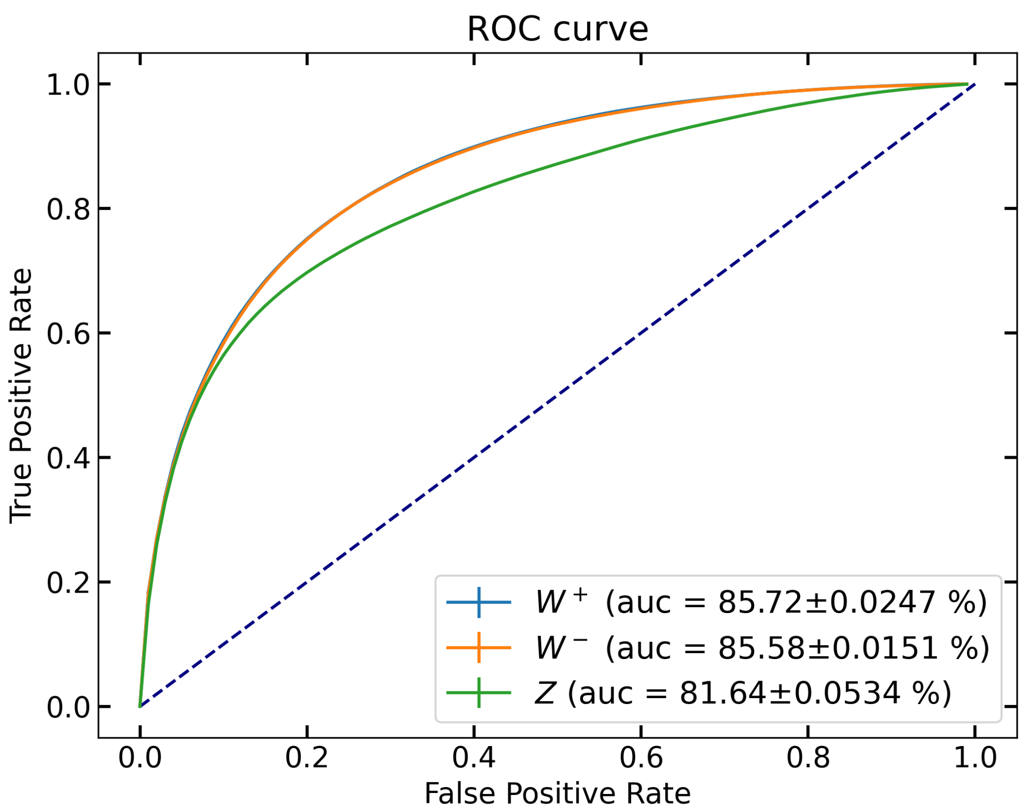



Figure 15: The ROC curve of each vector boson for the CNN model. The plotting scheme is the same as Figure 14.

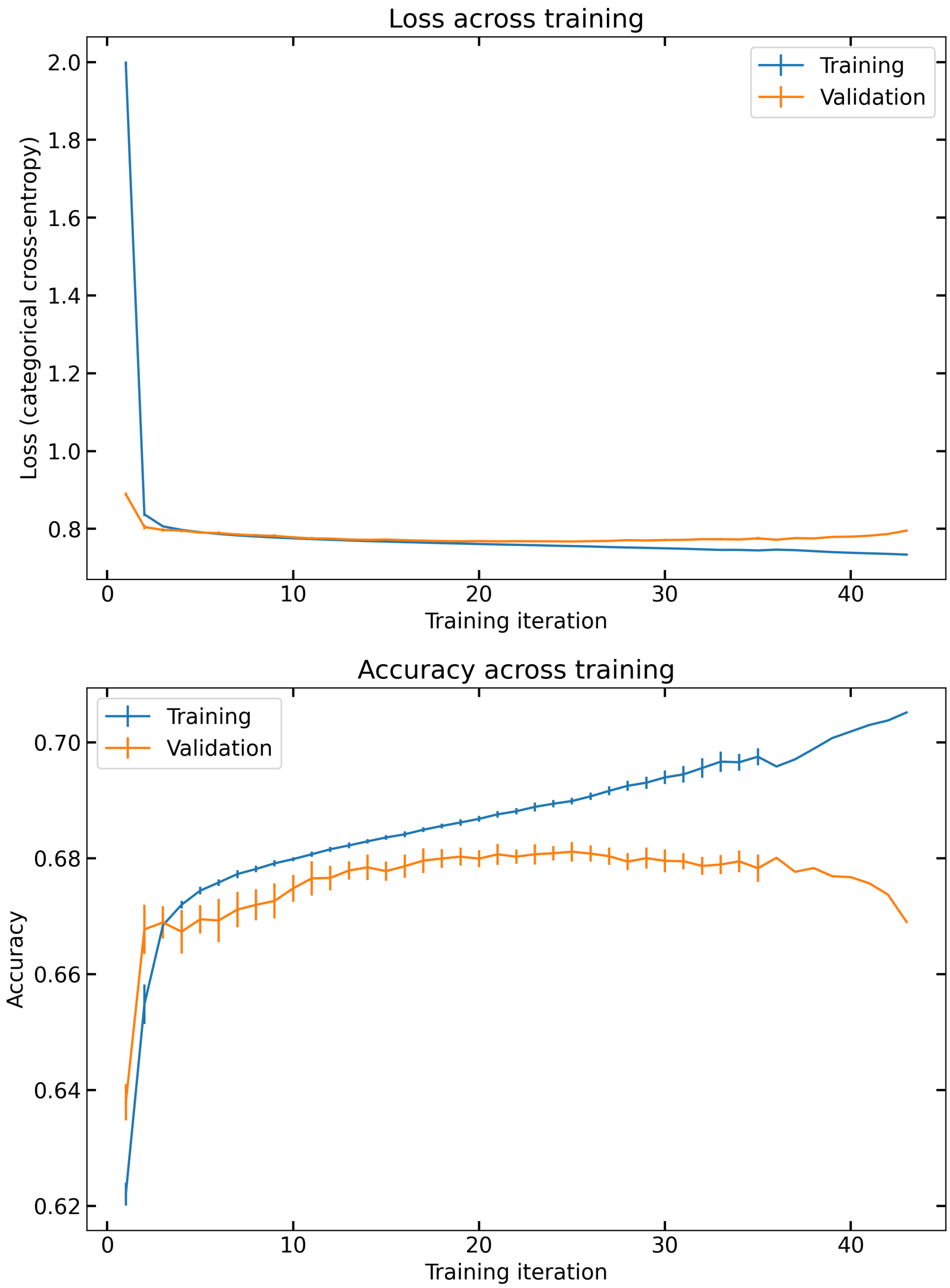



Figure 16: The loss and accuracy curve for the CNN$^2$ model. The plotting scheme is the same as Figure 14.

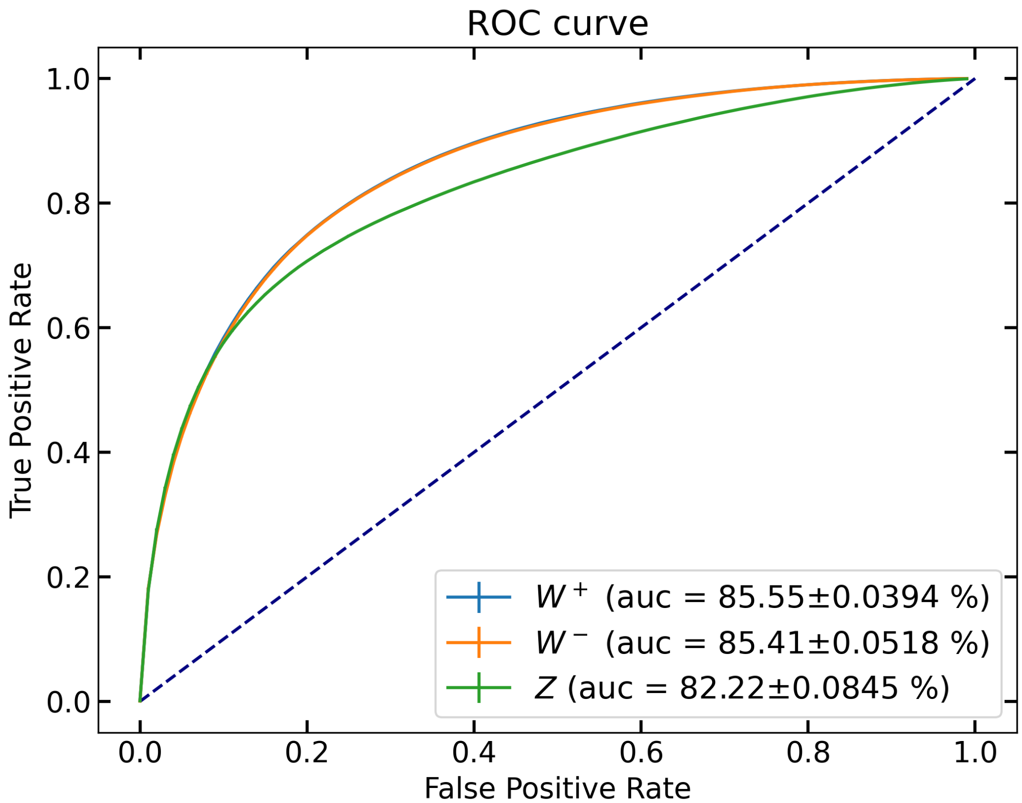
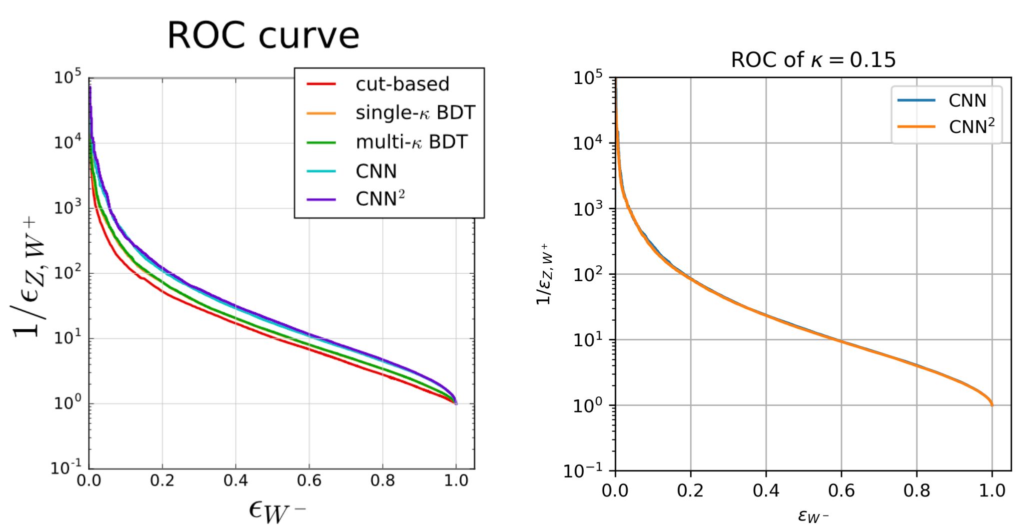


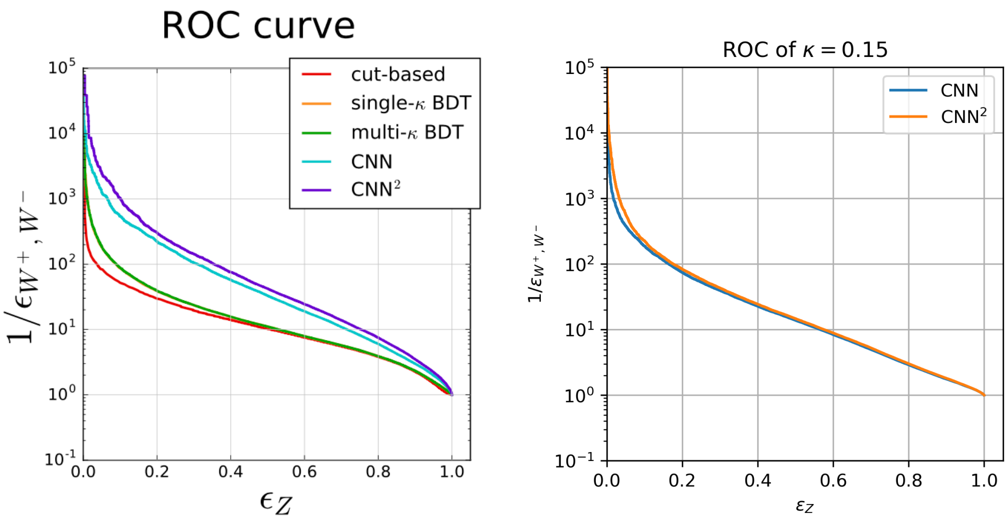

Figure 17: The ROC curve of each vector boson for the CNN$^2$ model. The plotting scheme is the same as Figure 14.




(a) Paper

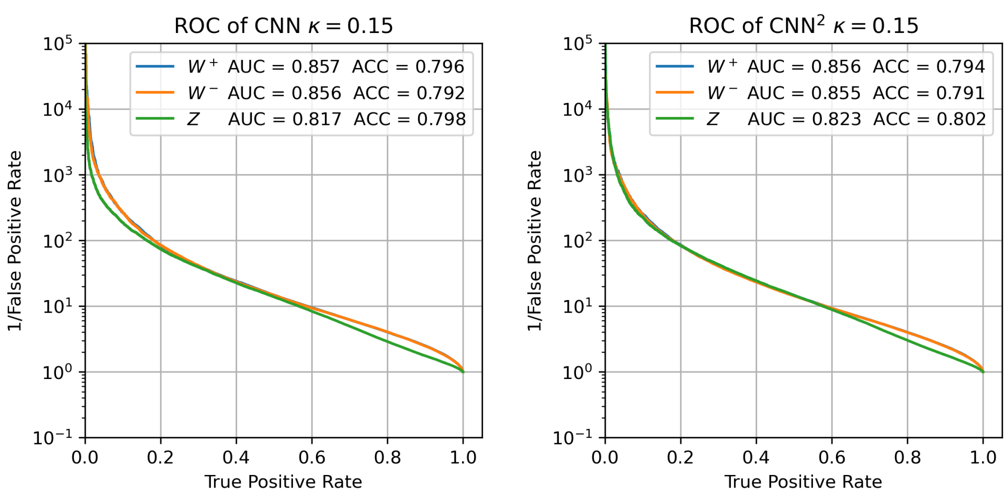
(b) Ours

Figure 18: The signal $W^-$ ROCs. Left is paper. Right is my result.




(a) Paper (b) Ours

Figure 19: The signal $Z$ ROCs. Left is paper. Right is my result.




(a) CNN (b) CNN$^2$

Figure 20: The ROCs of my training results. For the left one, the model is CNN. For right one, the model is CNN$^2$.